EXHIBIT A

Probabilistic determination of the body voltage in digital PD-SOI circuits and its application to static timing analysis

Abstract

We describe a technique for using knowledge of the average switching behavior of partially-depleted silicon-on-insulator (PD-SOI) circuits to reduce delay uncertainty produced by the floating-body effect. The approach is based on a unique state diagram abstraction of the PD-SOI FET which captures all of the essential device physics. A simple analytic model of the body voltage is derived and calibrated based on Monte Carlo simulations. This model is used within the context of a prototype transistor-level static timing analysis engine. Results are presented which demonstrate the dramatic reduction in delay uncertainty possible with this technique.

1 Introduction

Silicon-on-insulator (SOI) technology has long found niche applications for radiation-hardened or high-voltage integrated circuits. Recently, however, SOI has emerged as a technology for high-performance, low-power deep-submicron digital integrated circuits[1]. For digital applications, fully-depleted devices have been largely abandoned in favor of partially-depleted technology, because of the difficulty in controlling the threshold voltage of fully-depleted thin-film transistors. Partially-depleted SOI (PD-SOI) has two main advantages for digital applications: the reduction of parasitic source-drain depletion capacitances and the reduction of the reverse-body effect in stack structures and pass-transistor logic.

At the device and circuit level, however, the floating body effect in partially-depleted SOI (PD-SOI) poses major challenges in the successful use of this technology. There can be large "uncertainties" in the body potential based on past switching activities. For many circuits, the design margining required to protect against this uncertainty erodes all of the potential performance advantage under nominal operation. In addition, for many circuit styles in which noise margin is strongly determined by threshold voltage (e. g. dynamic circuits), considerable overdesign for noise can also result from conservative body-voltage margining.

Previous work on PD-SOI has focussed on device issues[2] or delay effects due to the floating-body effect evident for particular circuits under periodic stimulus[3, 4] (pulse stretching, frequency-dependent delay time). No technique has yet been developed for verifying floating-body effect on timing systematically across circuit styles. In this paper, we consider characterizing the body voltage uncertainty of PD-SOI devices using probabilistic knowledge of switching activity and applying this to transistor-level static timing analysis. In our approach, we analyze each FET of each circuit, determining the minimum (min) and maximum (max) body voltage under both "initial-condition" and "steady-state" operation. By initial-condition operation, we mean that the circuits are assumed to be quiescent for a long time and then undergo one cycle of switching activity to produce the min/max body voltage. Steady-state body voltage variation takes into account the average switching behavior of the circuit and is always considerably less conservative than the initial condition values. This technique formalizes the observation that one can expect very different behavior from a functional block on-chip that is quiescent for a long period of time as compared to one that has been under steady switching activity.

In this paper, we work with BSIM3SOI[5] model parameters for an IBM partially-depleted SOI technology described elsewhere[6]. Devices have a $0.25\mu m$ effective channel length, $5-nm$ gate oxide, $350-nm$ back oxide, and $140-nm$ thin silicon film. The supply voltage is 2.5V. While the detailed results we present here apply to this technology, the techniques are generally applicable to any PD-SOI technology.

In Section 2, we describe a state-diagram abstraction of the PD-SOI FET which simplifies the device physics determining the body voltage. In Section 3, we describe how we analyze this state diagram to determine the initial-condition and steady-state min/max body voltages. Section 4 describes a prototype transistor-level static timing analysis engine which incorporates these body voltage characterizations. Some results are presented in Section 5. Section 6 concludes and offers direction for future work.

2 State diagram view of body interactions

The body potential of a PD-SOI FET is determined by capacitive coupling of the body to the gate, source, and drain and by diode currents at the source-body and drain-body junctions. To model the switching history determining the current body voltage of a particular device, we use the state diagram abstraction shown in Figure 1 for an nFET. Arrows indicate possible state transitions produced by switching event, while the states themselves represent valid logic conditions for an FET at the end of a cycle or during hazards that occur transiently within a cycle. The states 1, 2, 5a, and 5b are distinguished in Figure 1 because transitions between these states propagate timing delays. States 5a and 5b can usually be treated equivalently as state 5. The state diagram of the pFET is the "dual" of Figure 1, in which the gate is high rather than low in states 3, 4, and 5; and low rather than high in states 1 and 2.

In analyzing the device physics determining the body potential, it is convenient to distinguish "slow" and "fast" processes. Fast processes can change the body potential on time scales on the order of or less than the cycle time, while slow processes require time scales much longer than the cycle time to affect the body voltage. There are two fast mechanisms at work: switching transitions on the gate, source, or drain which are capacitively coupled to the body, and forward-bias diode currents across source-body and drain-body functions with voltages exceeding the diode turn-on

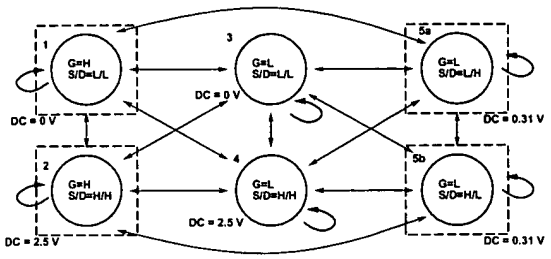

Figure 1: State diagram for a PD-SOI nFET.

voltage. The latter process results in a fast discharge of the body through the forward-biased junction. The coupling "kicks" associated with the first mechanism occur for each transition in Figure 1 and are completely reversible on "fast" time scales; that is, if one begins in state 1 and traverses the state diagram, returning to state 1 on a time scale comparable or less than the cycle time, the body voltage on return will be the same as the body voltage. The reversibility of the kicks is, of course, dependent on not triggering the irreversible discharge mechanism. In practice, the fast discharge mechanism is only triggered on activation of a circuit after a long period of quiescence. Once the discharge happens, it is not triggered again as long as the circuit is under steady switching activity.

The slow processes involve charging or discharging the body through reverse-biased or very weakly forward-biased diode junctions. These leakage currents give each state a slow relaxation to a target dc body voltage, which is indicated for the nFET case in Figure 1 for our target technology. We note that the dc body voltage of state 5 is determined by the steady-state balance between a weakly forward-biased junction and a reverse-biased junction. For the pFET case, the dc voltage of state 5 in our target technology is 1.73 V.

3 Determining the body potential

Important to determining the steady-state body voltage is characterizing the average switching behavior of the circuits. To do this, we appeal to stochastic techniques to determine the average fraction of time an FET spends in each of states 1 through 5. This is determined by two factors: the fraction of cycles an FET ends the cycle in a given state; and the amount of time the FET spends in each state with each cycle-to-cycle transition (including the possibility of hazards). We also find, based on simulation results, that to maximize (minimize) the body potential, we must favor those states with high (low) source-drain over those with low (high). In addition, although less importantly, we favor state 4(1) over state 2(3) for maximizing (minimizing) the body voltage for the nFET. For the pFET, we favor state 1(4) over state 3(2) for maximizing (minimizing) the body voltage. We expect these observations to extend generally to other PD-SOI technologies.

Each FET in our analysis is characterized by a set of arrival times and signal probabilities. For the nFET, these are:

- $P(G)$ = probability that at the end of a cycle the gate is high
- $P(D|\overline{G})$ = probability that at the end of a cycle the drain is driven high given that the gate is low
- $P(\overline{D}|\overline{G})$ = probability that at the end of a cycle the drain is driven low given that the gate is low
- $P(S|G) = P(D|G)$ = probability that at the end of a cycle the drain (and source) are driven high given that the gate is high
- $P(\overline{S}|G) = P(\overline{D}|G)$ = probability that at the end of a cycle that the drain (and source) are driven low given that the gate is low
- S^{early}_{rise}, S^{early}_{fall}, S^{late}_{rise}, S^{early}_{rise}, D^{early}_{rise}, D^{early}_{fall}, D^{late}_{rise}, D^{early}_{rise}, G^{early}_{rise}, G^{early}_{fall}, G^{late}_{rise}, G^{early}_{rise} = early and late arrival times (rising and falling) for the source, gate, and drain

By driven low (high), we mean that there is a path to ground (supply). The details of how these probabilities and arrival times are determined is described in Section 4. We note that $P(D|\overline{G}) + P(\overline{D}|\overline{G}) \leq 1$, for example, but does not have to sum to 1, since the conditional signal probabilities do not cover the case that the drain is floating.

Given these probabilities, the state diagram in Figure 1 represents a Markov process with two six-by-six transition matrices $\mathbf{A}$ (with one transition per machine cycle) corresponding to the min and max cases: $(P^i_{max})_{k+1} = \mathbf{A}_{max}(P^i_{max})_k$ and $(P^i_{min})_{k+1} = \mathbf{A}_{min}(P^i_{min})_k$. Each column of $\mathbf{A}_{max/min}$ is given by

$$(p_1^{max/min}\ p_2^{max/min}\ p_3^{max/min}\ p_4^{max/min}\ p_{5a}^{max/min}\ p_{5b}^{max/min})^T.$$

p_i is the probability of making a transition to state i. P_i is the probability of being in state i. For example, $p_3^{max} = (1 - P(G))P(\overline{S}|\overline{G})P(\overline{D}|\overline{G})$ while $p_3^{min} = (1 - P(G))(1 - P(\overline{S}|G))(1 - P(\overline{D}|G))$. The max (min) case assumes that the floating node condition goes high (low). Diagonalizing A (trivially) and finding the eigenvector associated with eigenvalue 1 (normalized so that the sum of the elements of the vector is 1) gives the steady-state values of the P_i. From these probabilities, one can calculate a set of thirty-six transition probabilities $P_{i\rightarrow j}$ that define the probability that a FET is current in state i transitioning to state j: $P_{i\rightarrow j} = P_i p_j$.

For each transition, we determine the fraction of the cycle time ($t_i^{j\rightarrow k}$) that can be spent in each state i as part of the transition $j \rightarrow k$ to maximize (or minimize) the body voltage. For example, if the nFET is transitioning from state 3 to state 3, then to maximize the body voltage:

$$t_4^{3\rightarrow 3} = Max(D^{late}_{fall}, S^{late}_{fall}) - Min(D^{early}_{rise}, S^{early}_{rise})$$
$$t_3^{3\rightarrow 3} = t_{cycle} - t_4^{3\rightarrow 3}$$

where t_{cycle} is the cycle time and the source and drain are not constant 0 and, therefore, can execute such a hazard during a cycle. From these times, we can calculate an effective amount of time (t_i^{eff}) on the average per cycle that the FET is in the state i:

$$t_i^{eff} = \sum_{j,k} P_{j\rightarrow k} t_i^{j\rightarrow k}$$

We now seek an analytic expression that allows us to determine the steady-state voltage from these t_i^{eff}.

If we imagine an average cycle broken into slices by t_i^{eff}, the body voltage in this average cycle will appear as in Figure 2. The discontinuities between the slices occur as a result

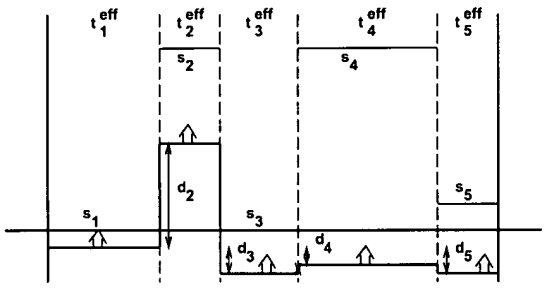

Figure 2: An average cycle.

of the capacitive coupling kicks and are characterized by the voltage differences d_i from the reference level (in this case, for the nFET, the body voltage of state 1). In addition, each state is relaxing (very slowly, which is why any voltage change during the time slice is imperceptible in Figure 2) to the target dc values noted in Figure 1 and here denoted as s_i. The time constants for this relaxation are denoted as τ_i. From this simple picture, one can relate the body voltage at the end of the cycle v_{n+1} to the body voltage at the beginning of the cycle v_n by:

$$v_{n+1} = v_n e^{\sum_{i=1}^{5} t_i/\tau_i} + \sum_{i=1}^{5} (s_i - v_i)(1 - e^{-t_i/\tau_i}) e^{\sum_{j=i}^{5} t_j/\tau_j}$$

The steady-state solution of this difference equation (in the approximatioin that the τ_i are much greater than the t_i^{eff}) is given by:

$$\frac{\sum_i (t_i/\tau_i)(s_i - d_i)}{\sum_i t_i/\tau_i}$$

We treat the relaxation time constants τ_i as fitting parameters that we determine to match the steady-state body voltages that result from Monte Carlo simulations of the state diagram in Figure 1. One such simulation is shown in Figure 3 for state 1 of the nFET and state 2 of the pFET. In this case the cycle time is 5 ns, all of the early arrival times are set to 1 ns, all of the late arrival times are set to 4 ns, and all of the signal probabilities are 0.5. In Figure 4, to demonstrate the efficacy of the analytic body voltage model, we compare it with Monte Carlo simulation results for an nFET (min and max). The conditions are the same as for Figure 3, except the signal probabilility of the gate is varied from 0 to 1.

In the case that the circuits can be quiescent for a long period of time before undergoing switching activity, one must calculate the minimum and maximum body initial-condition body voltages for states 1, 2, and 5. This calculation is done assuming that the device achieves a dc bias at one state and a capacitive kick (on a transition) into the target state. Based on simulation results for our target technology, we find that to maximize the body voltage, the initial state should be chosen with the priority: 4, 2, 5, 3, 1. Lower priority initial states are chosen only if high priority ones are probabilistically inaccessible. For the pFET, the initial state priority for the max case is: 2, 4, 5, 1, 3. The min case priority is the inverse of the max case priority. These initial conditions are noted for the example of Figure 3. The rapid progression from this state near the start of the simulation is due to forward-bias discharge of the body (a fast process). Slow processes alone then determine the steady-state. The steady-state value is independent of the the initial condition.

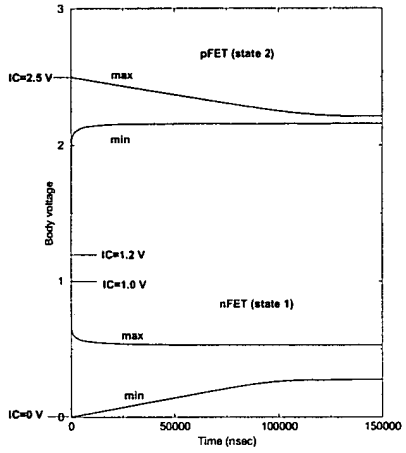

Figure 3: Body voltage transient.

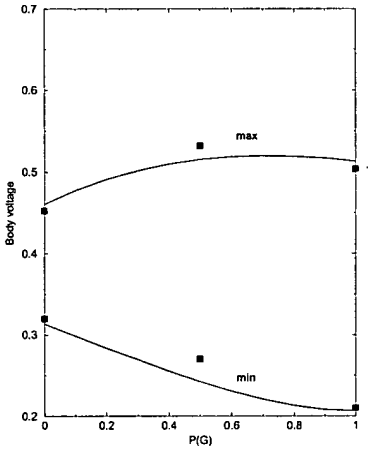

Figure 4: Body voltage fit.

4 Static timing analysis

We have incorporated this body voltage characterization into a prototype static transistor-level timing analysis engine utilizing a breadth-first-search algorithm. The design is partitioned into channel-connected components (CCC) for analysis, as is traditionally done in static transistor-level tools. The primary distinguishing feature of this timing analysis engine is that it propagates switching probabilities using assumptions of spatial and temporal independence, borrowing from similar techniques in static power analysis[7]. Once the signal probabilities and times are known at the inputs

of a CCC, these probabilities are translated into FET signal probability and arrival time properties for body-voltage calculation using an exhaustive path trace through the transistors of the CCC from the target node to supply or ground.

5 Results

Because of the length constraints of this paper, we present only two timing examples. The first is a 15-stage inverter chain. In Figure 5(a), we present the steady-state delay (for both rising and falling transitions) for each stage in the case that the switching probability at the input is 0.5. The minimum and maximum delays are virtually identical. We constrast this with Figure 5(b), in which we present the steady-state delay for the case that the switching probability at the input is 0.9. In this case, a wide disparity in the rising and falling delays opens up. In the case of the inverter, the only states accessible to the nFET are 1 and 5a, and for the pFET, 2 and 5b. For the inverters with input probabilities of 0.9, the steady-state body voltages of the pFET and nFET are pulled below their 0.5 values. This makes the nFET faster and pFET slower. The opposite applied to the inverters with input probabilies of 0.1 in the chain. The result is the well-known "pulse-stretching" phenomenon[4], now accessible within the static timing analysis environment.

We next consider an static eight-bit ripple carry adder using the full-adder implementation of Reference [8]. We consider the critical path of this circuit, the carry chain. In Figure 6(a), we show the stage delay using the initial-condition values of the body voltage. The minimum and maximum rising and falling delays are noted. The very large delay uncertainty is dramatically reduced by introducing knowledge that the circuit is under steady switching activity, as shown in Figure 6(b), which uses the steady-state analysis and assumes that all of the inputs have signal probabilities of 0.5. There is very little difference between the min and max cases in steady-state (very little delay uncertainty).

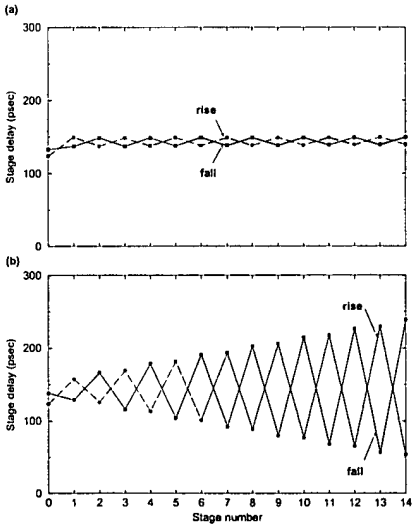

Figure 5: Inverter chain.

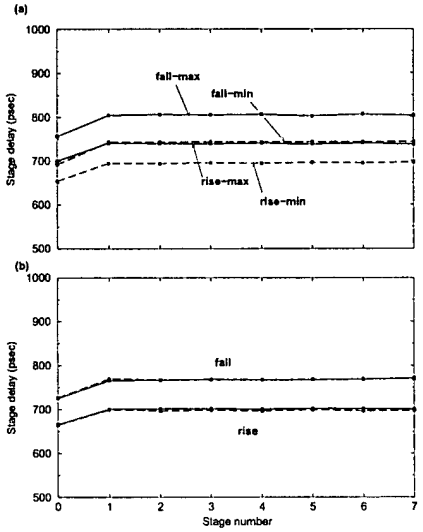

Figure 6: Carry chain.

6 Conclusions

We have demonstrated an efficient technique for bounding delay uncertainty in digital PD-SOI circuits using knowledge of the average switching behavior of the circuits, incorporating "hysteric" effects into the context of static timing analysis. Designers will naturally be very uncomfortable at first with applying stochastic techniques to something as "life-and-death" as timing or noise analysis. What they will find, however, is that the body-voltage uncertainty can be significantly reduced with fairly conservative assumptions about switching behavior. In many cases, a normally inactive block may be periodically stimulated to keep it "primed" so that when it is eventually exercised, it has more tightly controlled body voltage variation. Determining the necessary frequency and nature of this pattern and how it could be incorporated practically into the design remains an open research question.

References

[1] C. T. Chuang, P.-F. Lu, and C. J. Anderson. SOI for digital CMOS VLSI: Design considerations and advances. *Proceedings of the IEEE*, 86:689 – 720, 1998.

[2] J. Gautier and J. Y.-C. Sun. On the transient operation of partially depleted SOI NMOSFET's. *IEEE Electron Device Letters*, 16:497 – 499, 1995.

[3] R. Puri and C. T. Chuang. Hysteresis effect in pass-transistor-based partially-depleted SOI CMOS circuits. In *Proceedings of the International SOI Conference*, 1998.

[4] A. Wei, D. A. Antoniadis, and L. A. Bair. Minimizing floating-body-induced threshold voltage variation in partially depleted SOI CMOS. *IEEE Electron Device Letters*, 17:391 – 394, 1996.

[5] BSIM3SOI Manual. Technical report, University of California, Berkeley, 1998. Version 1.3.

[6] G. G. Shahidi et al. SOI for 1-volt CMOS technology and applictaion to a 512kb SRAM with 3.5 ns access time. In *Proceedings of the IEDM*, pages 813 – 816, 1993.

[7] Farid N. Najm. A survey of power estimation techniques in vlsi circuits. *IEEE Transations on VLSI Systems*, 2:446 – 455, 1994.

[8] R. K. Montoye, P. W. Cook, E. Hokenek, and R. P. Havreluk. An 18 ns 56-bit multiply-adder circuit. In *Proceedings of the IEEE International Solid-State Circuits Conference*, page 46, 1990.